



HAL
open science

Food, cooking and health in a selected corpus of websites and connected YouTube channels in France. Collecting and archiving the audiovisual web

Christian Bonah, Solène Lellinger, Caroline Sala

► To cite this version:

Christian Bonah, Solène Lellinger, Caroline Sala. Food, cooking and health in a selected corpus of websites and connected YouTube channels in France. Collecting and archiving the audiovisual web. Sophie Gebeil; Jean-Christophe Peyssard. Exploring the Archived Web during a Highly Transformative Age, 138 (1), Firenze University Press, pp.277-294, 2024, Proceedings e report, 979-12-215-0412-5. 10.36253/979-12-215-0413-2.24 . halshs-04704440

HAL Id: halshs-04704440

<https://shs.hal.science/halshs-04704440v1>

Submitted on 20 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Food, cooking, and health in a selected corpus of websites and connected YouTube channels in France. Collecting and archiving the audiovisual web

Christian Bonah, Solène Lellinger, Caroline Sala

Abstract: Based on a collaborative effort between the research project BodyCapital and the Bibliothèque Nationale de France (BnF), we present a two-step archiving process and analysis of audiovisual web content related to food and health history, investigating how audiovisuals have contributed to shaping our eating habits. The first step involved a web crawl with Heritrix, targeting 158 identified seed URLs compiled based on BnF science & technology lists and URLs identified by the research group. The crawl harvested 1,067,159 URLs. A content analysis identified 1,718 videos in our corpus. Content mapping and the identification of links to YouTube videos were performed, leading to the second step involving a focused collection of 34 YouTube channels harvesting 24,427 videos (2.4 TB) to be analyzed..

Keywords: audiovisuals, web archive videos, health, food history, YouTube.

Introduction

In her concluding chapter of the *Sage Handbook of Web History*, Jane Winters issues two calls. Firstly, the author advocates for a promising future “in which many different types of historian, not just those with an interest in contemporary politics or digital methods, can integrate web archives into their research” (Winters 2019, 596). Furthermore, Jane Winters argues that web archives and born-digital data might require or engender a radical reframing of avenues for historical research. She proposes a shift from focusing solely on textual elements of the archived web to incorporating a more comprehensive analysis of sound, still and moving images evermore present on the web. In doing so, historians “might engage with the history of art and design, media and communication studies, the history of technology, linguistics, film studies” (Winters 2019, 600).

However, historians working with web archives often experience disappointment when accessing reconstructed web pages that display the message “The video content cannot be read”. This is undoubtedly one of the reasons why the above-mentioned handbook includes a chapter on the sonic web (Wage Morris 2019), but not for web audiovisuals, digital-born or not. Yet audiovisual web content has become significant since the advent of

Christian Bonah, University of Strasbourg, France, bonah@unistra.fr, 0000-0003-4756-1844
Solène Lellinger, University of Paris, France, solene.lellinger@u-paris.fr, 0000-0002-9384-7360
Caroline Sala, University of Strasbourg, France, csala@unistra.fr, 0009-0007-1730-2135

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Christian Bonah, Solène Lellinger, Caroline Sala, *Food, cooking and health in a selected corpus of websites and connected YouTube channels in France. Collecting and archiving the audiovisual web*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0413-2.24, in Sophie Gebeil, Jean-Christophe Peyssard (edited by), *Exploring the Archived Web during a Highly Transformative Age. Proceedings of the 5th international RESAW conference, Marseille, June 2024*, pp. 277-294, 2024, published by Firenze University Press, ISBN 979-12-215-0413-2, DOI 10.36253/979-12-215-0413-2

photo-sharing tools like Flickr (in 2004) and video-sharing sites such as Vimeo (in 2004) and YouTube (in 2005).

Our research project lies at the intersection of working with web archives without being web historians, a specific interest in audiovisual content in web archives, and the intention to include web audiovisuals in a long-term audiovisual history spanning from traditional film to TV to the web. Our contribution will present a collaborative effort between the ERC research project BodyCapital at the University of Strasbourg and the Bibliothèque Nationale de France (BnF). It involves a two-step process aimed at archiving the audiovisual web related to food history and providing contextualization. Originating from health history and grappling with the methodological challenges of working from and writing through audiovisuals as central archival material, our research question seeks to understand how audiovisuals have contributed to shaping our eating habits and their connection to concerns about our individual health and healthy eating¹.

Methodological questions setting the stage: History through audiovisuals

Why audiovisuals? Exploring an entangled visual history (Lepenies 2003; Werner and Zimmermann 2004; Werner and Zimmermann 2006), our BodyCapital project investigates the impact of modern mass media, and especially visuals, throughout the twentieth century. Our focus is on understanding the role these media play in transforming our bodies into a form of capital and shaping individual receptiveness to the economization of health, reaching a point where individuals internalize the adoption of specific health practices and devices. In return, health practices and health-related goods emerge as a particularly stable and valuable vantage point from which to address twentieth-century changes in health conceptions and practices, national health policies and politics, and liberalizing market economies in Europe.

Three images from three audiovisuals in three media periods—the film *Mechanics of the brain* (Pudovkin, Russia, 1926), the TV program *Localisations cérébrales* (Jacques Rutman, France ORTF, 1964) and the internet video ego-document of American graduate student Ian Eslick (United States, 2011)—illustrate our point. Taken from a Russian educational documentary, the first image shows an experimental subject in Ivan Pavlov’s research laboratory with a primitive brain wave detector, focusing on the localization and analysis of the central nervous system. The second image, from a French TV film designed for teaching, presents a

¹ To illustrate the chapter, additional visuals (screenshots, data visualization, etc.) and audiovisual materials are available at: https://medfilm.unistra.fr/wiki/Corp:Food_cooking_and_health

patient with an electroencephalography headset in a hospital where doctors study the central nervous system for diagnosis and treatment purposes. The third one, sourced from an ego-document, shows an American student facing the camera in a staged shot. The quantifying self-activist registers his brain activity using an electrical sensor headband connected to a bedside base-station alarm clock (the commercialized Zeo REM sleep monitor). These three visuals suggest avenues for historical analysis of the incorporation of commodification, hybridization, evaluation, and the internalization of body techniques and devices, into individual, autonomous health practices throughout the twentieth century. The images indicate the fluctuating trajectories of visuals across social spaces (laboratories, schools, and the public sphere), different national spaces (Russia, France, and the United States), and different media formats with distinct aesthetics (experimental film, educational film, and internet video) over time (1926, 1964, and 2011). The three visuals juxtaposed here are evidently not equivalent in their original (visual) context. The first one was projected on a large screen, the second on a small school television set, and the third can be consulted on a computer or a smartphone. The body as capital in this comparison of visual sources shifts from exploitation to transformation and then self-investment, suggesting the analytical and heuristic potential of our audiovisual archive-based approach. Beyond their integration in visual sequences in time and space, images require contextualization through non-visual materials to situate their use and utility.

We conceive audiovisuals not merely as reflections or expressions of observed phenomena, but as media forms possessing their own distinct, interactive, and performative power. We consider them essential for several key reasons. Firstly, their distribution has experienced considerable expansion; secondly they transcend professional and social groups, thirdly because of their utilitarian character, and fourthly their complementarity with economic market principles enhances their role in the promotion/communication. More explicitly, they are essential for our approach, which seeks to track and understand changes of bodily health perceptions and practices, since (1) visuals at the age of mechanical reproduction (Benjamin 1935) have become missionary communication tools extending public relations to hard-to-reach segments of society, such as illiterates, rural populations, and lower classes (see, for example, lantern shows and early cinema as fair attractions and genuinely popular culture) (Gunning 1990; Strauven 2006; Elsaesser 2009). (2) Visuals have traversed social worlds and time, evolving from clinical and laboratory settings to various forms such as health education brochures, posters, films, television programs, and video. They have even found a place in consumer goods, transforming into communication tools like X-ray images printed on cigarette packages to convey cancer risks. (3) New (audio)visual media—

from motion pictures to television and the internet—have been revolutionary technologies throughout the twentieth century. They have entertained, documented, instructed, and transformed mass audiences, similar to how moveable type transformed medieval Europe (McLuhan and Lapham 1994). Yet, visuals are more than an ideal vantage point for research; they represent conditions and conditioning that have transformed word-based health politics—akin to knowledge and scientific objectivity with and without words (Daston and Galison 2007)—into less verbal and more visual communication vectors essential to the enactment of health beliefs and practices in the twentieth-century communication society. Communication involving visuals requires methodological and analytical reflection. This involves considering how visuals, beyond their symbolic functioning from pictograms to smileys, relate to pre-existing discourse, accompanying lectures or interpretative words. Analysis of visual and non-visual material is therefore needed to account for the fact that visuals in social contexts are generally associated with or interpreted by printed, screened, or spoken words. (4) Since the interwar period, visuals have been conceived as indispensable tools for the “invisible government” (Bernays 1928), acting as an alter ego to the “invisible hand” of the market and taking the form of promotion-communication and corporate public relations.

Building on this commitment, our research project investigates how audiovisuals have changed body politics and influenced the self-perceptions and practices related to health in individuals within market-based societies in twentieth-century Europe. Following on from film and television, our contribution conveys a systematic audiovisual web analysis focusing on the subject of food studies health, eating, and nutrition.

Crawling the audiovisual web # 1: Creating a food, eating, nutrition and health web archive corpus

In collaboration with the Digital Legal Deposit Unit of the Bibliothèque nationale de France (BnF) and the Bibliothèque Nationale Universitaire de Strasbourg (BNUS), we initiated a web crawl using the Heritrix robot covering 158 identified seed URLs compiled based on BnF science and technology lists and URLs identified by the research group.² The crawl conducted between March 9, 2021 and March 12, 2021 harvested 1,067,159 URLs at a level $n+2$. A content analysis using SolrWayback identified 1,718 videos within our food, eating, nutrition, and health corpus (FENH).

² We are grateful to Alexandre Faye, Sara Aubry, Isabelle Degrange, and Leslie Bellony from the BnF, and Jérôme Schweitzer and Madeleine Hubert from the BNUS, for their help and guidance.

The BnF, and the BNUS in delegation, are two libraries in France that hold the responsibility of the legal deposit of the web. The rationale behind this collaboration was to benefit from their extensive experience of over ten years in collecting archives, their capacity to archive the data permanently, their proficiency in indexing and the availability of indexing tools, along with their capability to provide critical review with web archive specialists throughout the research process. Monthly work meetings covered remote access in Strasbourg, indexing advancements, problems encountered, and critical feedback from both parties. This rationale seemed especially useful for us as historians working with web archives without being web historians.

In more detail, the first crawl targeted French food-related websites, with a focus on scrutinizing representations of food on the web. This encompassed themes such as food and health, food producers and processing, cooking, agriculture, diets, and other relevant topics including forms of intermediality and audiovisuals. Seed URLs were selected from pre-existing lists of the BnF Science and Technologies Department, which has constructed a decade-long thematic collection. These were further validated by the expertise of researchers from the BodyCapital group. A few websites related to recent trends absent from the pre-existing list were added to enrich the selection. In total, 84 website homepages (primary URLs) were gathered, complemented by an additional 88 targeted tabs (secondary URLs). The complete list is recorded in the BnF curator tool called *Collecte du web* (BCWeb). A general description and keywords (topics covered, producer type, media included) were entered for each item, providing an overview of the entire selection. The crawl settings were defined as follows: Page +2, allowing the bot to explore content with a maximum depth of 50,000 URLs per website. The crawl, conducted by the BnF in March 2021, produced a relatively small but representative, controlled, and curated corpus. The uncompressed data size is 88.17 GiB and the WARC files are preserved by the BnF, with a copy delivered to the BNUS. In compliance with French legal deposit legislation, a dedicated workstation at the BNUS was made available for our use.

SolrWayback³ software was chosen as the tool to initiate the analysis of the specifically collected FENH web archive corpus. This allowed us to quickly and efficiently index the WARC files, including images and video, and to test numerous features provided by the tool. Our first analysis involved producing comprehensive textual and visual mind maps. These were organized by themes and image typologies, highlighting the use of

³ SolrWayback, developed by the Royal Library of Denmark, is quickly becoming an important tool for the archivist community. For further information about SolrWayback see: Anders Klindt Myrvoll's presentation of the tool: IIPC WAC 2022: SESSION 1 #2: SolrWayback at the Royal Danish Library https://youtu.be/-q4a-edVP5E?si=EmxMXrhpQ_wG4WY9.

archetypes in the representation of diets and food types. Our case study focused specifically on color and gesture. However, our video analysis encountered several challenges, including superficial content descriptions, difficulties in reading or retrieving videos for close visual analysis, struggles with diverse video file formats, and, notably, the absence of video content in the collected archives due to their hosting outside the website itself.

To address and complement the limitations of our fragmented video corpus produced through an URL-based crawl, we opted for a second crawl, once again with the BnF, but this time centered on YouTube channels initially identified by the first crawl and further complemented by lists compiled by web archiving specialists from the BnF. The supplementary YouTube crawl, the second step in our approach, was initiated 24 months after the initial URL crawl.

Crawling the audiovisual web # 2: Complementing the FENH corpus with a targeted YouTube channel crawl

YouTube hosts videos from all the content producers identified in step #1 (URL crawl), underscoring the importance of video communication for all players (companies, associations, and individuals). Following discussions with the BnF, we decided to initiate a second crawl focused on preserving YouTube channels related to the topic of food. Leveraging the BnF web archivists' experience crawling YouTube and our familiarity with the platform as researchers and users, we collected data from 34 French channels representing 24,427 videos (MP4 format: 2.4 TB) in June 2022, thus constituting our FENH video corpus. This second collection includes data for each channel retrieved from the YouTube API, stored in JSON format by the BnF. These data are also used to reconstruct a YouTube page needed for accessing the video collection through the BnF public application. A dedicated *Guided Tour* (*parcours guidé* in French) allows readers to browse all the preserved channels.

Our methodology for selecting YouTube channels resembles the approach to URL selection, combining information collected beforehand by the BnF web archivists and cross-referencing input from the BodyCapital team. Detailed channel descriptions were incorporated, allowing researchers (and the audience of the guided BnF presentation) to understand the nature of the content they encounter and why it was considered relevant to our research project. In collaboration with the BnF, we produced a detailed description of 36 harvested YouTube channels, including NGOs, national organizations, individual blogs, food processing professionals, and industrial actors. These descriptions are based on the analysis of channels and individual videos within each YouTube channel. Providing data on the

channel (name, author, content, goals) and highlighting the video's importance for the history of food and health, these descriptions also include links to websites and blogs harvested in step 1, or links to other social media platforms such as Instagram.⁴ Working with video files posed major challenges, including managing data volume and file size, the need for operating media readers, and the frequent absence of connected descriptions of the file content.

The research conducted on the URL crawl, taking into account the time lapse between the two crawls, allowed us to analyze data from step #1 and to complete the second YouTube crawl, particularly addressing missing types of actors or producers and issues of under- and over-representation.

Content providers range from industrial actors to culinary chefs, (health)food-related associations to cooking brands, and private blogs to agri-food institutions. The categorization of content producers was predetermined in accordance with BnF guidelines. In a preliminary analysis of channels to produce the highest number of videos, notable channels included *Marmiton* (2,465 videos, online recipes from a cooking brand); *Ptitchef.com* (1,122 videos, online recipes); *Chef Michel Dumas* (681 videos, professional chef); *Les fruits et légumes frais* (479 videos, health and nutrition association); *Chez Jigmé* (469 videos, recipes by an individual YouTuber). The most active channels predominantly focus on recipe postings. To prevent saturation of the corpus with dominant content such as recipes, the number of providers was limited. As a result, our corpus does not claim to represent the entire food video landscape on the web. Rather, our interest lies in the variety and multiplicity of video genres and types. Instead of studying hundreds of similar recipe videos with minimal variations beyond the dish prepared, our goal is to conduct a visual analysis of short-format videos, examining their directing style, settings, camera angles, protagonists, plots, intertwined textual elements, sound, and music. Ultimately, our analysis aims to yield a better understanding of their role as contemporary mass media in the internet era and highlight their similarities and differences with traditional, prevailing audiovisual media forms like film and television. At the same time, our research targets a long-term twentieth-century analysis of food habits, cooking practices, and perceptions of healthy nutrition. This extends into the larger process of transforming the act of nourishing our bodies into investing in our body capital, generating individual receptiveness to the economization of food

⁴ The YouTube channels involved in the project are described here:

https://www.bnf.fr/sites/default/files/2021-03/videos_parours.pdf. In the 'Internet Archive' consultation application, a facet has been introduced to limit the search to selections made as part of the research project. With this link: <http://archivesinternet.bnf.fr/parours/18-videos/projetrecherche/bodycapital>

and eating, and internalizing changing paradigms for the adoption of healthy food consumption.

The subsequent presentation will delve into some of the results of the two-step crawl, both in quantitative and qualitative terms, and will propose analytical approaches for the analysis of the collected corpus. Furthermore, the thematic analysis will investigate a comparison between the results of the two-step crawl and the audiovisual formats and themes of a television program corpus that has been systematically collected and analyzed in parallel. The trans-media analysis will provide insights into the transformations between television and the web, particularly regarding specific thematic sub-corpus.

Cross-sectional analysis of a two-step crawl: Content providers and opinion leaders, a case study

A first step towards analyzing our FENH corpus involved identifying highly visible content producers and their contributions. Given the overlap between the two crawls, our emphasis was on producers present in both collections. These producers were then grouped into several categories⁵ (Figure 1).

⁵ Aprifel – Agency for research & information on fruit & vegetables; Anses – French Agency for Food, Environmental and Occupational Health & Safety; Ameliore ta santé – Enhance your health; CrudiVegan – Raw and vegan food; Cuisine saine – Healthy food, gluten-free, lactose-free; Ptitchef.com – Recipes; Vitagora – Agri-food innovation cluster; Cultures Sucre – Association of French sugar beet growers and sugar manufacturers; Nestlé France; ANIA – French National Association of Food Industries; CERIN nutrition – Nutritional Resources and Information Center, health department of the dairy products interprofession; LaBananeInfo – Interprofessional Banana Association; Produits laitiers – Consumer-oriented information of French dairy interprofessional organization.

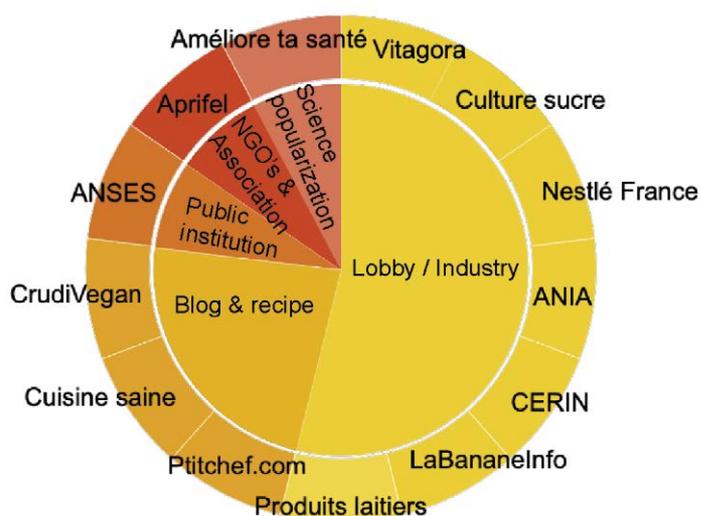


Figure 1 : Typology of content producers and their presence at the intersection of URL and YouTube channel crawls (presence in both crawls).

Five main categories of producers and contents emerge both from the collection of URLs and from the YouTube channels. Firstly, there is a group of producers from the agri-food industry, professional groups, and lobbies. Secondly, a major category revolves around cooking recipes, often originating as blogs and evolving into professional websites and media platforms. A third category concerns public institutions. A fourth category consolidates content produced by NGOs and associations, while the final category focuses on different forms of science popularization. It is important to note that this overview is not representative of the entire web, but is influenced by the collection of choices described above. Nevertheless, the overall distribution and typology indicate a substantial presence of agri-food industry players and media groups investing in food and cooking activities. Agri-food industries and nutrition/cooking web media account for 75% of the audiovisual media content in our FENH corpus. In some cases, such as *Marmiton*, the content producer began as a private website and blog for exchanging recipes and, due to its success, transformed into a professional web media entity acquired by media groups. At times, agri-food industries with longstanding commitments to audiovisual media communication expanded their activities to the new media landscape of the web in the late 2000s.

Our previous work on healthcare lobbying (Lellinger and Bonah 2021)—in particular, the commercial practices of the pharmaceutical industry—led us to select a specific case study to explore in greater detail. Our choice

concerns a producer from the agri-food category, more specifically the case of the dairy industry. One notable example is the YouTube videos produced by the French Dairy Interbranch organization (CNIEL), such as “Dairy products are our friends for life Mister V”, which garnered over 7 million views within its first five years on the web.⁶ Concurrently, the CNIEL has a longstanding commitment to public communication, with highly successful television spots promoting milk and dairy products dating back to the 1980s. Given our integrative long-term audiovisual approach, the “Les Produits Laitiers” (Dairy products) website and YouTube channel are of particular interest due to their extensive prehistory, including television, poster, and even film campaigns (Nourrisson 2020; Zimmermann 2021).⁷ The CNIEL’s public engagement with broad and specifically targeted audiences has been studied and the numerous television commercials produced since the 1980s promoting dairy products in France coined the famous and enduring slogan “Les produits laitiers sont nos amis pour la vie” (dairy products are our friends for life) (Nourrisson 2021). The CNIEL’s dairy production promotion through television spots in the 1980s invested and engaged with specific audiences. Animation and pop music catered to entertainment forms specifically aimed at the younger generations. Different spots addressed nutritional questions at a more scientific level, targeting adult audiences with specific expectations regarding healthy eating. These spots functioned as a hybrid of promotion and education hyphens, showcasing the CNIEL’s commitment to social responsibility for promoting milk and dairy products as integral components of a healthy diet. “Milky way didactics” incorporated incentives for “good consumerism”, emphasizing nutritional value and offering advice that intertwines food education with product promotion (Nourrisson 2002). Through web archives network analysis, we can track and analyze these trends following the shift of the CNIEL’s communication and promotion to the web sphere. Research into the CNIEL web organization and culture reveals an intricate constellation of websites affiliated with the dairy industry. The French Dairy Interbranch organization, CNIEL, made separate websites for each interaction, promotion, and lobbying theme structured around at least 16

⁶ *Produits Laitiers MISTER V - LES COPAINS AU LAIT - EP.1 LE LOTO*, 2021-06-06-1 145 308 views. Page archived on: 22 June 2022 at 19:50 GMT in the videos collection <http://archivesinternet.bnf.fr/20220622195046/http://www.youtube.com/watch?v=6A1qrXnrtWM>. Accessible from BnF Internet Archives.

⁷ For a comparative perspective: Tricia Close-Koenig’s unpublished work on milk and dairy product promotion in Canada, the USA, and the UK. Different films are available at MedFilm: *Three for Health* (1950, Canada, https://medfilm.unistra.fr/wiki/Three_for_Health), *The Milky Way* (1939, Canada, https://medfilm.unistra.fr/wiki/The_Milky_Way), *Grandpa’s Party* (Canada, https://medfilm.unistra.fr/wiki/Grandpa%27s_Party), *The Milky Way* (1948, UK, and MedFilm analysis at: https://medfilm.unistra.fr/wiki/The_Milky_Way_UK Accessed February 5th, 2024.

related websites with content tailored to specific objectives and diverse audiences.

A more detailed network analysis of the “Produits Laitiers” website and YouTube channel’s outgoing links confirms the CNIEL web sphere described above. The first 20 nodes of outgoing links from their website, generated with SolrWayback, notably highlights social media (YouTube, Facebook, Twitter, Instagram), cooking-oriented social media (Youmiam), traditional media (cheriefm.fr), public health-related websites (mangerbouger.fr, santepubliquefrance.fr), and professionally related websites of the French dairy interprofessional organization, and food industry associations. More surprisingly, the analysis also reveals connections with a museum institution (Quai Branly)⁸.

The positioning of *Produits-laitiers.com* as a significant player in food education is highlighted by its outlinks to state-controlled public health websites dedicated to health education, including *santepubliquefrance.fr* and *mangerbouger.fr*. The implied responsibility for health education aligns with dairy product promotion, turning health education into a persuasive argument for the milk industry and its web presence and publications. The *Produits-laitiers.com* website sits at the crossroads of public health issues, on the one hand, and professional promotion and lobbying on the other (Kratz 2024).

The word occurrence and association analysis based on page content (Figure 2) after cleaning our crawl data⁹ produces a wordcloud illustrating that the main topic categories are, of course, dairy products (milk, cheese, etc.), but subjects such as healthy eating, well-being, eco-responsibility/ethics, and considerations for children and teenagers also feature prominently. The centrality of well-being, balanced diet, and dairy products encapsulate the website's central message.

⁸ See the figure “Outgoing link analysis of *Produits-laitiers.com*. Based on the FENH corpus, SolrWayback” on the companion website:

https://medfilm.unistra.fr/wiki/Corp:Food_cooking_and_health

⁹ The cleanup consisted of removing linking words, certain HTML tags, and certain terms that generated noise, such as ‘Cookies’. We used JupyterNotebook for data analysis.

The YouTube channel content exhibits a diverse range of productions targeting various audiences, encompassing agriculture reports, cooking recipes, educational content, pure slapstick entertainment, and straightforward advertising. Health, education and social responsibility remain high on the *Produits-Laitiers* YouTube channel agenda. Societal concerns, such as lactose intolerance, are tackled head on. The few-minutes long video from 2014 “Lactose intolerance”, which mimics a purely educational format, reassures viewers that “90-95% of the French population can digest a bowl of milk without any problem” and that everyone produces the enzymes necessary for milk digestion. Educating and promoting dairy products within the context of societal questioning is one of the key video production lines of *Produits-Laitiers.com*.

Sophisticated entertainment remains the second pivotal axis for *Produits Laitiers*. Shifting promotional content to the web, the CNIEL has significantly invested in commercial and media collaborations facilitated by social networks, engaging highly visible journalists, influencers, and web videographers. The *Produits Laitiers* YouTube channel even reintroduces older successful television concepts, playing on the nostalgia of older-generation ‘webspectators’.

Illustrating this practice, the video “Les différents laits, présentés par Jamy” (Different forms of milks, presented by Jamy)¹¹ features Jamy Gourmand, a popular science journalist who gained recognition in the 1990s–2000s with his TV show “It’s not rocket science” (*C’est pas sorcier*), broadcasted on the public television channel *France 3*. The concept of the broadcast involved duplexing from his science truck, with journalists Frédéric Courant and Sabine Quindou, explaining science to children. In 2020, Jamy Gourmand established his own YouTube channel, *Epicurieux*, dedicated to science popularization, currently boasting over 1.7 million subscribers. In October 2021, the renowned journalist produced a video for the *Produits Laitiers* YouTube channel, explaining the different forms of milk, leveraging both his past TV presentations and his recent involvement in YouTube video production.

The CNIEL website ecosystem finds its audiovisual translation in an imitation of the reputed Netflix video streaming platform, aptly named “100% milk videos”. Playing on words and slogans once again, *Lait’Flix X* hosts dairy promotion videos grouped under headings like entertainment, nutrition & health, agriculture and animal well-being, or sports activities and milk. *Lait’Flix* incorporates humorous videos from YouTubers, such as

¹¹ *Produits Laitiers : Les différents laits, présentés par Jamy*, 2021-10-28-363 242 views, Page archived on 22 juin 2022 at 19:50 GMT in the videos collection <http://archivesinternet.bnf.fr/20220622195058/http://www.youtube.com/watch?v=6nTPXX72JkQ>
Accessible from BnF Internet Archives.

one featuring the YouTuber *Mister V*, who boasts 6.3 million subscribers, or offers cooking recipes centered around... dairy products.

Back to the web archives at the BnF

The primary and summary analysis of our two-step audiovisual web crawl and the CNIEL case study suggests several directions for further investigation. On the one hand, we could broaden and deepen the analysis in terms of the collection scope whether spatial and/or temporal. It appears beneficial to study the interconnections between the various CNIEL sites in greater detail, investigating audience-specific translations operating from one *Produits Laitiers* website to another. This would facilitate the analysis of the thematic environment specific to each information and communication space.

As demonstrated by the analysis of outgoing links, we could, by extension, study the textual and especially audiovisual contents of sites to which the *produits-laitiers.com* site links, beyond the related CNIEL sites.

Another avenue of exploration is of a more ‘archeological’ nature. BnF legal deposit web archive collections allow us to go back in time. The historical depth of BnF website archives, collected over many years, permits consultation of archives from 2006. At that time, *Produits Laitiers* had already initiated its web presence and featured the “le Blog des Produits Laitiers”.

Original, sophisticated, and fostering consumer bonding, the *Produits Laitiers* website had been offering Flash content, including games for both children and adults, since 2010. While this type of content has now disappeared from the current web, it underscores the ongoing commitment of the CNIEL to pioneering promotional practices in the most recent media spaces.

Pitched to teenagers and school children, the “three milk powers (growth, strength, and balance)” present challenges to be met. 3D films, quizzes, interactive games, and puzzles captivate first-generation web users. Once again, entertainment and health education intertwine when the headline announces a “challenge in between sensory pleasure and health”.

Web sphere ecologies can be traced over time through outgoing link analysis. Studying referrals from the *Produits-laitiers.com* website to other sites in 2015, for example, indicates referrals made to the Parisian spring attraction *Salon de l'Agriculture*, an annual agriculture show in Paris. This referral also features the 2021 outgoing link analysis presented above.

The historical depth and interconnections explored in this way involve examining audiovisual content and beyond to understand how it adapted to

the communications challenges and expectations of different eras and the ‘trends’ of the moment, such as blogging, flash games¹², and streaming.

For the historian, our case study highlights the ability to move back and forth between different types of archives. It allows historians to link them together, extending beyond the simple web archive consultation, thereby reconstructing long-term dynamics in promotional-educational practices that condition consumer behavior at a lifelong generational level. The methodological choice of using a public institution for web legal deposit not only guarantees the continuity of the collection, but also empowers historians to navigate between various archival collections, be they web archives or archives of prior audiovisual media.

Conclusion

Our two-step web crawl-based approach, assembling a food, eating, nutrition, and health web video corpus (FENH corpus), represents an effort to systematically analyze audiovisual sources on the web and their archives. This approach aims to overcome challenges posed by the often-difficult accessibility and analysis of audiovisual web content. Collaborating with the BnF as a legal deposit institution not only ensures the continuity of data hosting, but also provides opportunities to (1) transition from data to visual studies of archived material; (2) investigate genealogies of audiovisuals on the web and beyond-before; (3) shift from outgoing link analysis to relational audiovisual content analysis; and eventually (4) address ephemeral audiovisual web content.

The *Produits Laitiers* case study inevitably prompts the question of whether the case represents more than itself. Given its extensive size and distribution of the French web, coupled with the fact that 75% of audiovisual media content in our FENH corpus originates from agri-food industries and nutrition and cooking web media, the analysis raises the question of whether the commercial web has significantly marginalized public actors, NGOs, and associations, and if so, when did this occur? Was the utopian egalitarian vision of the early web merely a chimera from the beginning, as the *Produits Laitiers* analysis may suggest? In any case, the personalized content approach of web promotion catering to individual users has led to a multiplication of audience-specific targeting. Interactive strategies, involving forms to be filled out for inquiries, advice, or exchanges, point in the same direction. This direction becomes evident in our wordcloud analysis, which names specific audiences (mothers, teenagers, user match, etc.)

¹² See, for example, this site, which offers flash games and a challenge for schools. <http://archivesinternet.bnf.fr/20100422153616/http://www.les3pouvoirslaitierstedefient.com>

In our broader research goal to integrate web audiovisual productions into a century-long history written through successive and overlapping audiovisual media forms, our analysis positions web audiovisuals as a third-era source for long-term studies of the role of audiovisuals in shaping our eating habits. A transmedial and intermedial approach opens up new avenues for *long-durée* research based on cross-analysis of audiovisual formats from archives. These formats portray and question continuities and changes in dairy promotion over a century. From *The Milky Way* (1948), a central galaxy of food products in Western modernity, agriculture, and their promotion to *Milk products are our friends for life* (1981), agricultural lobbying and food industries have consistently invested in their product promotion. Perhaps part of their success lies precisely in their generational long-term and life-long promotional-educational impact.

Milk films and videos have spanned the 20th century, evolving from depictions of milk processing and distribution (*Milky Way*, UK, 1948) to portraying milk as a lifestyle dancing companion (*Milk products our friends for life*, TV spot, 1981), and eventually to comical media self-reflection and consumer persiflage (*Lait'Flix: Mister V*, web video, 2021/2019). From school lunch campaigns in the 1950s to critical appraisals in the 2010s, the image of the milk and dairy products has undergone significant changes and challenges (Souccar 2008). However, milk promotional campaigns have endured throughout the century, transitioning from arguments about reconstructing the nation after WWII to promoting personal nutritional emancipation and self-enhancement through milk consumption. Adopting recent video and YouTube trends, web archives reveal the continued engagement of milk promotional efforts in the third-media era, adapting to changes in media and society.

References

- Benjamin, Walter. 1935. *The Work of Art in the Age of Mechanical Reproduction*. <https://web.mit.edu/allanmc/www/benjamin.pdf>.
- Bernays, Edward. 1928. *Propaganda*. New York: Horace Liveright.
- Daston, Lorraine, and Peter Louis Galison. 2007. *Objectivity*. New York: Zone Books.
- Elsaesser, Thomas. 2009. "Archives and Archaeologies. The Place of Non-Fiction Film in Contemporary Media." In *Films That Work: Industrial Film and the Productivity of Media*, edited by Vinzenz Hediger and Patrick Vonderau, 19–34. Film Culture in Transition. Amsterdam: Amsterdam University Press.
- Gunning, Tom. 1990. "The Cinema of Attractions. Early Film, Its Spectator and Avant-Garde." In *Early Cinema: Space, Frame, Narrative*, edited by Thomas Elsaesser and Adam Barker, 56–63. London: BFI Pub.
- Kratz, Amélie. 2024. "Au-Delà de l'éducation Nutritionnelle : L'éducation Au Manger Des Enfants à Travers Les Audiovisuels En France et En République Fédérale d'Allemagne (1950–1980)." PhD thesis, University of Strasbourg.
- Lellinger, Solène, and Christian Bonah. 2021. "'This Corporation Has "Anesthetized" the Actors in the Drug Chain'. Influence Peddling and the Normality of Conflicts of Interest in the Mediator® Scandal." In *Conflict of Interest and Medicine*, 181–200. London: Routledge.
- Lepenies, Wolf, ed. 2003. *Entangled Histories and Negotiated Universals: Centers and Peripheries in a Changing World*. Frankfurt: Campus Verlag.
- McLuhan, Marshall, and Lewis Henry Lapham. 1994. *Understanding Media the Extensions of Man*. Cambridge (Mass.): The MIT Press.
- Nourrisson, Didier, ed. 2002. *A votre santé !: éducation et santé sous la IVe République*. Saint-Etienne, France: Publications de l'Université de Saint-Etienne.
- Nourrisson, Didier. 2020. "Faire-savoir et savoir-faire à l'École : l'alimentation à l'écran." *Les Enjeux de l'information et de la communication* 20/3A (S1): 117–130. doi:<https://doi.org/10.3917/enic.hs9.0117>.
- Nourrisson, Didier. 2021. *Du lait et des hommes: histoire d'un breuvage nourricier de la Renaissance à nos jours*. Paris: Éditions Vendémiaire.
- Souccar, Thierry. 2008. *Lait, mensonges et propagande*. Vergèze: T. Souccar.
- Strauven, Wanda, ed. 2006. *The Cinema of Attractions Reloaded. Film Culture in Transition*. Amsterdam: Amsterdam University Press.
- Wage Morris, Jeremy. 2019. "Hearing the Past: The Sonic Web from MIDI to Music Streaming." In *The SAGE Handbook of Web History*, edited by Niels Brügger and Ian

- Milligan. London; Thousand Oaks, California: SAGE Publications.
- Werner, Michael, and Bénédicte Zimmermann, eds. 2004. *De la comparaison à l'histoire croisée*. Paris: Seuil.
- Werner, Michael, and Bénédicte Zimmermann. 2006. "Beyond Comparison: Histoire Croisée and the Challenge of Reflexivity." *History and Theory* 45 (1): 30–50.
- Winters, Jane. 2019. "Web Archives and (Digital) History: A Troubled Past and a Promising Future?" In *The SAGE Handbook of Web History*, edited by Niels Brügger and Ian Milligan. London; Thousand Oaks, California: SAGE Publications.
- Zimmermann, Yvonne. 2021. "Early Cinema, Process Films, and Screen Advertising." In *Advertising and the Transformation of Screen Cultures*, edited by Bo Florin, Patrick Vonderau, and Yvonne Zimmermann, 21–48. Amsterdam: Amsterdam University Press.